CANONICAL FORMS:

NOTES FOR RESEQUENCING PARTS OF A MATRIX ALGEBRA COURSE

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, New York 14853

BU-1220-M September, 1993

ABSTRACT

Newly-written course notes for Statistics 417, "Matrix Algebra", replace Chapter 7 and include parts of Chapters 11 and 11A of the course text *Matrix Algebra Useful for Statistics.* These notes improve the course in the following ways.

(a) The important topic of eigen roots and vectors is introduced mid-course rather than at the end. This introduction deals only with symmetric matrices since, for statistics, the most important canonical form is that under orthogonal similarity of a symmetric matrix. Some details and proofs are left until the end of the course. Instead, emphasis is placed on the existence and uses of the orthogonal similar form, and on development therefrom of other forms for symmetric matrices (e.g., congruent and diagonal forms, and spectral decomposition).

(b) The equivalent canonical form of any (real, non-null) matrix is developed solely from row operations; and the messy arithmetic of the text's Chapter 7 can be trashed — which should have happened years ago. Rather than that, emphasis is placed on the existence and uses of $\mathbf{PAQ} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ without ever needing numerical values of $\mathbf{P}$ and $\mathbf{Q}$.

(c) Through introducing eigen roots and vectors only for symmetric matrices, not only is that important class of matrices dealt with mid-course, but also the somewhat complicated diagonability theorem can be left until the end of the course.

Grateful acknowledgment goes to C. E. McCulloch for initially motivating these notes.

CANONICAL FORMS:

NOTES FOR RESEQUENCING PARTS OF A MATRIX ALGEBRA COURSE

PART I: SYMMETRIC MATRICES 2

1. An introductory example 2
   (a) Eigen roots (e-roots) 3
   (b) Eigen vectors (e-vectors) 4
   (c) A canonical form 5
2. Canonical form under orthogonal similarity: eigen roots and vectors 6
   (a) General properties 7
   (b) Example 8
   (c) Quadratic forms 9
3. Other factorizations 10
   (a) The congruent canonical form and the equivalent canonical form 10
   (b) Full rank factorization 12
   (c) Diagonal forms 13
   (d) Spectral decomposition 13
   (e) Quadratic forms 14
   (f) Non-negative definite matrices 14
      –i. Eigen roots are non-negative 14
      –ii. Positive definite matrices have positive eigen roots 14
      –iii. Congruent canonical forms 15
      –iv. Full-rank factorization 15

PART II: ROW OPERATIONS ON ANY MATRIX 16

4. Row operators 16
   (a) Adding a multiple of a row to a row 16
   (b) Interchanging two rows: permutation matrices 17
   (c) Multiplying a row by a constant 17
5. Rank and row operators 18
   (a) Square matrices 18
   (b) Rectangular matrices 20
6. Diagonal forms 21
   (a) An example 21
   (b) The general case 23
   (c) Conclusion 24
7. The equivalent canonical form 25
8. Using the equivalent canonical form 25
   (a) Full rank factorization 25
   (b) Non-singular matrices 27
   (c) Rank of a product matrix 27
9. The $Q - R$ factorization 29
   (a) Example: a square matrix 30
   (b) Rectangular matrices – more rows than columns 31
   (c) Inverting $\mathbf{A} = \mathbf{QR}$ 31

CANONICAL FORMS PART I: SYMMETRIC MATRICES

We deal here with the broad topic of canonical forms. For a square matrix $\mathbf{A}$, a general description of a canonical form is that it is a diagonal matrix (with some or none of its diagonal elements zero) that can be obtained by pre- and post-multiplication of $\mathbf{A}$ by certain other matrices. For three very good reasons we introduce the topic by first dealing with canonical forms of symmetric matrices. The reasons are as follows:

(i) Matrices that are symmetric are the matrices that occur most frequently in statistics; e.g., displaying variances of, and covariances between, a set of n random variables leads to what is called a variance-covariance matrix (see text, p. 347), which is symmetric; and equations that often arise in statistics are $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ and $\mathbf{X}'\mathbf{X}$ is, of course, symmetric.

(ii) Canonical forms of symmetric matrices are easier to deal with than are those of non-symmetric matrices.

(iii) The most useful and important canonical form of a symmetric matrix is that known as the canonical form under orthogonal similarity. We therefore begin with this, starting with an example that is a simple, symmetric matrix of order 2.

1. AN INTRODUCTORY EXAMPLE

A symmetric (Markov) matrix that arises in studies of the human genome and elsewhere is

$$\mathbf{F} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}, \tag{1}$$

where $p < 1$ is a probability. $\mathbf{F}$ occurs where its successive powers are often of interest, e.g.,

$$\mathbf{F}^2 = \begin{bmatrix} 1-2p+2p^2 & 2p-2p^2 \\ 2p-2p^2 & 1-2p+2p^2 \end{bmatrix}$$

and (2)

$$\mathbf{F}^3 = \begin{bmatrix} 1-3p+6p^2-4p^3 & 3p-6p^2+4p^3 \\ 3p-6p^2+4p^3 & 1-3p+6p^2-4p^3 \end{bmatrix},$$

and so on, as may be obtained by successive multiplication. But note that not only is $\mathbf{F} = \mathbf{F}'$, but also

$\mathbf{F1} = \mathbf{1} = \mathbf{F}^2\mathbf{1} = \mathbf{F}^3\mathbf{1}$ and, indeed, $\mathbf{F}^k\mathbf{1} = \mathbf{1}$ for all positive and negative integers k (so long as $p \neq 1$).

The equality $\mathbf{F1} = \mathbf{1}$ prompts us to ask the question "More generally than $\mathbf{F1} = (1)\mathbf{1}$, is there a vector $\mathbf{u}$ and a scalar λ such that

$$\mathbf{Fu} = \lambda\mathbf{u}\ ?\text{"} \tag{3}$$

The answer is "Yes". Moreover, after finding $\mathbf{u}$ and λ explicitly there is a much simpler way of calculating powers of $\mathbf{F}$ than using the successive multiplications that yielded $\mathbf{F}^2$ and $\mathbf{F}^3$ in (2).

a. Eigen roots (e-roots)

It is clear that (3) can be rewritten as

$$(\mathbf{F} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0}\ . \tag{4}$$

Now if $\mathbf{F} - \lambda\mathbf{I}$ is non-singular its columns are LIN. But the left-hand side of (4) represents a linear combination of columns of $\mathbf{F} - \lambda\mathbf{I}$. Hence the only way for (4) to hold when $\mathbf{F} - \lambda\mathbf{I}$ is non-singular is for $\mathbf{u}$ to be null, $\mathbf{u} = \mathbf{0}$. And that is of no interest in (3) because (3) is *always* true for $\mathbf{u} = \mathbf{0}$. Therefore we investigate $\mathbf{F} - \lambda\mathbf{I}$ being singular, in which case

$$|\mathbf{F} - \lambda\mathbf{I}| = 0\ , \tag{5}$$

i.e.,

$$\left|\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right| = 0\ .$$

Hence

$$\begin{vmatrix} 1-p-\lambda & p \\ p & 1-p-\lambda \end{vmatrix} = 0$$

and so

$$(1-p-\lambda)^2 - p^2 = 0\ ,$$

which means

$$1-p-\lambda = \pm\, p\ .$$

Thus

$$\lambda = 1 \text{ or } 1-2p$$

and we denote these two values as

$$\lambda_1 = 1 \qquad \text{and} \qquad \lambda_2 = 1-2p\ . \tag{6}$$

So now we have two values of λ for which it appears (3) can exist. These values of λ, which satisfy

(5), are called the *eigen roots* of **F**. (Other names are eigen values, characteristic roots or values and, sometimes, latent roots or values.) For brevity's sake we will call the eigen roots "*e-roots*". The equation (5) is called the *characteristic equation* of **F**.

b. Eigen vectors (e-vectors)

Having found e-roots λ for **F**, two of them because **F** has order 2, we now seek in (3) a value of **u** corresponding to each *e-root*. That corresponding vector is called an *eigen vector* (or characteristic or latent) *vector*. We abbreviate this to *e-vector*.

Corresponding to λ_1 and λ_2 we obtain these vectors by successively using λ_1 and λ_2 of (6), in (3), denoting elements of $\mathbf{u}_i$ for λ_i, by α_i and β_i. For $\lambda_1 = 1$, equation (3) is

$$\begin{bmatrix} 1-p-1 & p \\ p & 1-p-1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix} = \mathbf{0}$$

i.e.,

$$-p\alpha_1 + p\beta_1 = 0$$
$$p\alpha_1 - p\beta_1 = 0 \; .$$

Clearly a solution is $\beta_1 = \alpha_1$, i.e.,

$$\mathbf{u}_1 = \begin{bmatrix} \alpha_1 \\ \alpha_1 \end{bmatrix}$$

and this is for *any* value of α_1, so to begin with we take $\alpha_1 = 1$ and get

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} .$$

For $\lambda_2 = 1-2p$, equation (3) is

$$\begin{bmatrix} 1-p-(1-2p) & p \\ p & 1-p-(1-2p) \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \beta_2 \end{bmatrix} = \mathbf{0} ,$$

which is

$$p\alpha_2 + p\beta_2 = 0$$
$$p\alpha_2 + p\beta_2 = 0$$

with $\beta_2 = -\alpha_2$ as a solution, i.e.,

$$\mathbf{u}_2 = \begin{bmatrix} \alpha_2 \\ -\alpha_2 \end{bmatrix} .$$

This is for any value of α_2, and so we take $\alpha_2 = 1$ as a start:

$$\mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

There are many properties of λ_1, λ_2 and $\mathbf{u}_1$, $\mathbf{u}_2$ that could now be discussed because as e-roots and e-vectors of a symmetric matrix they have numerous properties of interest. But we defer details of these to the discussion of the general case that follows, noting here some of these properties for the example.

(c) A canonical form

First observe that

$$\mathbf{u}_1' \mathbf{u}_2 = [1 \quad 1] \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0 \;;$$

i.e., that the e-vectors are orthogonal. Let us normalize them (see top of text page 70), but retain the same symbols for them so that they become

$$\mathbf{u}_1 = \frac{1}{\sqrt{1^2+1^2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \frac{1}{\sqrt{1^2+(-1)^2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{7}$$

[The two norms, $\sqrt{2}$, are the same solely because of the nature of $\mathbf{F}$; that equality is not generally true.]

With the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ in (7) retaining their orthogonality we now have

$$\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2] = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{8}$$

as an orthogonal matrix:

$$\mathbf{U}\mathbf{U}' = \mathbf{U}'\mathbf{U} = \mathbf{I}\,.$$

Also

$$\mathbf{F}\mathbf{u}_1 = (1)\mathbf{u}_1 \qquad \text{and} \qquad \mathbf{F}\mathbf{u}_2 = (1-2p)\mathbf{u}_2\,,$$

which can be written in one equation as

$$\mathbf{F}[\mathbf{u}_1 \quad \mathbf{u}_2] = \left[1(\mathbf{u}_1) \quad (1-2p)\mathbf{u}_2\right] = [\mathbf{u}_1 \quad \mathbf{u}_2] \begin{bmatrix} 1 & 0 \\ 0 & 1-2p \end{bmatrix}. \tag{9}$$

On writing $\mathbf{D}$ as

$$\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 1-2p \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{10}$$

we then have (9) as

$$\mathbf{F}\mathbf{U} = \mathbf{U}\mathbf{D}\,, \tag{11}$$

and so, because **U** is orthogonal,

$$\mathbf{U'FU} = \mathbf{D} \qquad \text{with} \qquad \mathbf{U'U} = \mathbf{I} = \mathbf{UU'} \ . \tag{12}$$

This is known as the *canonical form under orthogonal similarity* of the symmetric matrix **F**. The comparable result for a non-symmetric matrix is discussed in the text, Section 11.5.

The result in (12) is very useful. For example, it easily gives powers of **F**. From (12)

$$\mathbf{F} = \mathbf{UDU'}$$

and so

$$\mathbf{F}^2 = \mathbf{UDU'UDU'} = \mathbf{UDIDU'} = \mathbf{UD}^2\mathbf{U'} \ .$$

Similarly

$$\mathbf{F}^3 = \mathbf{UD}^3\mathbf{U'}$$

and

$$\mathbf{F}^k = \mathbf{UD}^k\mathbf{U'} \ ,$$

and with **D** being diagonal $\mathbf{D}^k$ is easy to calculate. Thus on substituting from (8) and (10)

$$\begin{aligned} \mathbf{F}^k &= \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1^k & 0 \\ 0 & (1-2p)^k \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1+(1-2p)^k & 1-(1-2p)^k \\ 1-(1-2p)^k & 1+(1-2p)^k \end{bmatrix} . \end{aligned} \tag{13}$$

An important consequence of (13) is that for this particular matrix **F**, (13) easily provides the limiting value of $\mathbf{F}^k$ as k tends to infinity:

$$\lim_{k \to \infty} \mathbf{F}^k = \mathbf{I} \ .$$

2. CANONICAL FORM UNDER ORTHOGONAL SIMILARITY: EIGEN ROOTS AND VECTORS

The procedure detailed for **F** of (1) that led to the canonical form under orthogonal similarity, namely $\mathbf{F} = \mathbf{UDU'}$ of (12), is simply a special case of a general result that applies to any symmetric matrix, whatever its order, n say. (As usual, we confine attention to real matrices.)

(a) General properties

We here state a few of the general properties, leaving proofs, and applicable generalizations to non-symmetric matrices, until Chapter 11.

(i) For $\mathbf{A} = \mathbf{A}'$, an e-root and e-vector, λ and $\mathbf{u}$, respectively, satisfy $\mathbf{Au} = \lambda\mathbf{u}$.

(ii) $|\mathbf{A} - \lambda\mathbf{I}| = 0$ is the characteristic equation of $\mathbf{A}$.

(iii) That equation has n roots, $\lambda_1, \cdots, \lambda_n$; they are the e-roots of $\mathbf{A}$.

(iv) Some of those e-roots may be equal to one another, including the possibility that one or more are zero.

(v) Because $\mathbf{A}$ is symmetric, none of the e-roots is a complex number; every one of them is real.

(vi) Every e-root λ_i has an e-vector $\mathbf{u}_i$ associated with it; and if some value occurs k times as an e-root, there will be k e-vectors corresponding to that value. (This statement applies to all symmetric matrices but only to some non-symmetric matrices. The diagonability theorem in Chapters 11 and 11a deals with this difficulty.)

(vii) When $\mathbf{A}$ is symmetric, all the n e-vectors are orthogonal to each other, and each can be normalized. They can then be arrayed as the columns of a matrix $\mathbf{U}$ and, just as in (9) for $\mathbf{F}$, we can have, for $\mathbf{A} = \mathbf{A}'$,

$$\mathbf{A}[\mathbf{u}_1 \ \ \mathbf{u}_2 \ \ \cdots \ \ \mathbf{u}_n] = [\mathbf{u}_1 \ \ \mathbf{u}_2 \ \ \cdots \ \ \mathbf{u}_n]\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \tag{14}$$

which is

$$\mathbf{AU} = \mathbf{UD} \tag{15}$$

for $\mathbf{D}$ being a diagonal matrix of all n e-roots. $\mathbf{U}$ is orthogonal, because

$$\mathbf{A} = \mathbf{A}', \ \text{ and so } \ \mathbf{A} = \mathbf{UDU}' \text{ and } \mathbf{U}'\mathbf{AU} = \mathbf{D}\ . \tag{16}$$

This is the *canonical form under orthogonal similarity.*

In (14) and (15) note that the labelling of the e-vectors from 1 to n in $\mathbf{U}$ must correspond exactly to the labelling of the e-roots λ_i in the diagonal matrix $\mathbf{D}$.

(viii) So saying, it is customary to put all the zero e-roots in the lower right-hand corner of $\mathbf{D}$ so that it can be partitioned as

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{17}$$

where $\mathbf{D}_r$ is diagonal, containing the non-zero e-roots of $\mathbf{A}$.

(ix) For symmetric $\mathbf{A}$

$$\text{order of } \mathbf{D}_r = \text{number of non-zero e-roots} = \mathrm{r}(\mathbf{A}) .$$

(x) The equality $\mathbf{Au} = \lambda\mathbf{u}$ leads to

$$(\mathrm{c}\mathbf{A})\mathbf{u} = (\mathrm{c}\lambda)\mathbf{u} \quad \text{and} \quad \mathbf{A}(\mathrm{c}\mathbf{u}) = \lambda(\mathrm{c}\mathbf{u})$$

for any scalar c. These equalities show that when λ is an e-root of $\mathbf{A}$ with corresponding e-vector $\mathbf{u}$, then $\mathrm{c}\lambda$ is an e-root of $\mathrm{c}\mathbf{A}$ corresponding to that same e-vector $\mathbf{u}$; and for e-root λ of $\mathbf{A}$, a corresponding e-vector is not only $\mathbf{u}$, but also $\mathrm{c}\mathbf{u}$.

b. Example The characteristic equation for

$$\mathbf{A} = \mathbf{A}' = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \quad \text{is} \quad \begin{vmatrix} 1-\lambda & 2 & 2 \\ 2 & 1-\lambda & 2 \\ 2 & 2 & 1-\lambda \end{vmatrix} = 0 \tag{18}$$

which reduces to

$$(\lambda-5)(\lambda+1)^2 = 0 \quad \text{with roots } \lambda_1 = 5 \text{ and } \lambda_2 = -1 = \lambda_3 .$$

For $\lambda_1 = 5$, the equation $(\mathbf{A}-\lambda\mathbf{I})\mathbf{u}_1 = \mathbf{0}$ is

$$\begin{aligned} -4\alpha_1 + 2\beta_1 + 2\gamma_1 &= 0 \\ 2\alpha_1 - 4\beta_1 + 2\gamma_1 &= 0 \\ 2\alpha_1 + 2\beta_1 - 4\gamma_1 &= 0 \end{aligned} \quad \text{giving} \quad \mathbf{u}_1 = \begin{bmatrix} \alpha_1 \\ \alpha_1 \\ \alpha_1 \end{bmatrix} \quad \text{for any } \alpha_1 .$$

For $\lambda_2 = -1$, the equation $(\mathbf{A}-\lambda_2\mathbf{I})\mathbf{u}_2 = 0$ is

$$\begin{aligned} 2\alpha_2 + 2\beta_2 + 2\gamma_2 &= 0 \\ 2\alpha_2 + 2\beta_2 + 2\gamma_2 &= 0 \\ 2\alpha_2 + 2\beta_2 + 2\gamma_2 &= 0 \end{aligned} \quad \text{giving} \quad \mathbf{u}_2 = \begin{bmatrix} -(\alpha_2+\beta_2) \\ \alpha_2 \\ \beta_2 \end{bmatrix} \quad \text{for any } \alpha_2 \text{ and } \beta_2 .$$

Assigning numerical values to α_1, α_2 and β_2 we take

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2^* = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix};$$

and we notice immediately that these three vectors are orthogonal. To assemble them in an orthogonal matrix $\mathbf{U}$, we first normalize them, using divisors $\sqrt{1^2+1^2+1^2} = \sqrt{3}$, $\sqrt{(-2)^2+1^2+1^2} = \sqrt{6}$ and $\sqrt{0^2+1^2+1^2} = \sqrt{2}$, respectively, so that the normalized $\mathbf{u}_i$ are

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2^* = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}.$$

Then, on using the common denominator $\sqrt{6}$, $\mathbf{U}$ is

$$\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_2^*\end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix}\sqrt{2} & -2 & 0\\ \sqrt{2} & 1 & -\sqrt{3}\\ \sqrt{2} & 1 & \sqrt{3}\end{bmatrix}. \tag{19}$$

The reader should verify that

$$\mathbf{U'AU} = \mathbf{D} = \begin{bmatrix}5 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & -1\end{bmatrix}, \tag{20}$$

the diagonal matrix of e-roots with the k'th column of $\mathbf{U}$ being the e-vector corresponding to the k'th diagonal element in $\mathbf{D}$.

Note: The simple arithmetic involved in calculating the e-vectors for this example and for the earlier $\mathbf{F}$ does not extend easily to all matrices. A general methodology given in Section 11.4 of the text will be dealt with in class in due time.

(c) Quadratic forms

The canonical form under orthogonal similarity as just derived is useful in enabling any quadratic form $\mathbf{x'Ax}$ (with $\mathbf{A}$ always being taken as symmetric – see text page 75) as a sum of squares each with a + or − sign. This is so because, with $\mathbf{U'AU} = \mathbf{D}$

$$\begin{aligned}\mathbf{x'Ax} &= \mathbf{x'UU'AUU'x}, \qquad \text{because } \mathbf{U} \text{ is orthogonal}\\ &= \mathbf{x'U(U'AU)Ux}\\ &= \mathbf{y'Dy} \qquad \text{with} \qquad \mathbf{y} = \mathbf{U'x}\\ &= \sum_{i=1}^{r}\lambda_i y_i^2\ ,\end{aligned}$$

which is a sum of squares $\left(\sqrt{\lambda_i}y_i\right)^2$ where λ_i is real, but can be positive or negative.

In the preceding example, $\mathbf{U}$ of (19) gives $\mathbf{y} = \mathbf{U'x}$ as

$$\mathbf{y} = \frac{1}{\sqrt{6}}\begin{bmatrix}\sqrt{2} & \sqrt{2} & \sqrt{2}\\ -2 & 1 & 1\\ 0 & -\sqrt{3} & \sqrt{3}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_2\end{bmatrix} = \begin{bmatrix}(x_1+x_2+x_3)\big/\sqrt{3}\\ (-2x_1+x_2+x_3)\big/\sqrt{6}\\ (-x_2+x_3)\big/\sqrt{2}\end{bmatrix},$$

whereupon on using $\mathbf{D}$ of (20)

$$\begin{aligned}\mathbf{x'Ax} &= \mathbf{y'Dy}\\ &= 5\left[(x_1+x_2+x_3)\big/\sqrt{3}\right]^2 + (-1)\left[(-2x_1+x_2+x_3)\big/\sqrt{6}\right]^2 + (-1)\left[(-x_2+x_3)\big/\sqrt{2}\right]^2\end{aligned}$$

which expands out to be

$$\mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 + x_2^2 + x_3^2 + 4x_1x_2 + 4x_1x_3 + 4x_2x_3 \,,$$

obtained by using **A** directly, as in (30) on page 75 of the text.

3. OTHER FACTORIZATIONS

From (16), we have $\mathbf{A} = \mathbf{UDU}'$. This can be modified in a variety of ways that are useful in many different situations.

(a) The congruent canonical form and the equivalent canonical form

A particularly useful result comes from writing

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

of (17), where $\mathbf{D}_r$ is a diagonal matrix with its diagonal elements being the r non-zero e-roots of **A**. Because those roots are real [see (ix) above], there exists a matrix $\mathbf{\Delta}$ say, such that $\mathbf{\Delta}^2 = \mathbf{D}_r$, i.e., $\mathbf{\Delta}$ is diagonal with diagonal elements that are the square roots of those of $\mathbf{D}_r$; and so we write $\mathbf{\Delta} = \sqrt{\mathbf{D}_r}$. Then for

$$\mathbf{W} = \begin{bmatrix} \sqrt{\mathbf{D}_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{21}$$

$$\mathbf{A} = \mathbf{UWW}^{-1}\mathbf{DW}'^{-1}\mathbf{W}'\mathbf{U}' \tag{22}$$

$$= \mathbf{UW}\begin{bmatrix} \left(\sqrt{\mathbf{D}_r}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \left(\sqrt{\mathbf{D}_r}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\mathbf{W}'\mathbf{U}' \tag{23}$$

$$= \mathbf{Q}\begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{Q}'$$

for $\mathbf{Q} = \mathbf{UW}$, non-singular. And so for $\mathbf{P} = \mathbf{Q}^{-1}$

$$\mathbf{PAP}' = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{24}$$

The matrix on the right-hand side of (24) is known as the *canonical form under congruence*, of the symmetric matrix **A**. The comparable result for non-symmetric matrices, including rectangular matrices, is called the *equivalent canonical form* as in (47).

The importance of (24) is that it *always* exists, for any symmetric matrix of rank r. In general, calculating **P** is tedious; and there are many such matrices **P** since any particular **P** depends upon the order in which the non-zero e-roots are arrayed in $\mathbf{D}_r$. But we are seldom interested in knowing a specific value for **P**; what is so useful is the fact that for any symmetric matrix there is *always* a non-singular **P** such that the canonical form under congruence, (24), exists. Despite our seldom needing **P** we show an example of calculating it.

Example

$$\mathbf{A} = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & -2 \\ 0 & -2 & 2 \end{bmatrix}.$$

It will be found that **A** has e-roots of 3, 6 and 0, and that $\mathbf{A} = \mathbf{UDU}'$ is

$$\mathbf{A} = \tfrac{1}{3}\begin{bmatrix} 2 & 2 & 1 \\ 1 & -2 & 2 \\ -2 & 1 & 2 \end{bmatrix}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix}\tfrac{1}{3}\begin{bmatrix} 2 & 1 & -2 \\ 2 & -2 & 1 \\ 1 & 2 & 2 \end{bmatrix}$$

and so (23) is

$$\mathbf{A} = \mathbf{UW}\begin{bmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{6} & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{6} & 0 \\ 0 & 0 & 1 \end{bmatrix}\mathbf{W}'\mathbf{U}'$$

$$= \mathbf{Q}\begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{Q}'$$

and

$$\mathbf{Q} = \mathbf{UW} = \tfrac{1}{3}\begin{bmatrix} 2 & 2 & 1 \\ 1 & -2 & 2 \\ -2 & 1 & 2 \end{bmatrix}\begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{6} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \tfrac{1}{3}\begin{bmatrix} 2\sqrt{3} & 2\sqrt{6} & 1 \\ \sqrt{3} & -2\sqrt{6} & 2 \\ -2\sqrt{3} & \sqrt{6} & 2 \end{bmatrix}$$

$$\mathbf{P} = \mathbf{Q}^{-1} = (\mathbf{UW})^{-1} = \mathbf{W}^{-1}\mathbf{U}^{-1} = \mathbf{W}^{-1}\mathbf{U}'$$

$$= \begin{bmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{6} & 0 \\ 0 & 0 & 1 \end{bmatrix}\tfrac{1}{3}\begin{bmatrix} 2 & 1 & -2 \\ 2 & -2 & 1 \\ 1 & 2 & 2 \end{bmatrix} = \tfrac{1}{3}\begin{bmatrix} 2/\sqrt{3} & 1/\sqrt{3} & -2/\sqrt{3} \\ 2/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \\ 1 & 2 & 2 \end{bmatrix}.$$

The reader should confirm these results by calculating (24); and should notice that if

$$\mathbf{D}_r = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \quad \text{had been written as} \quad \begin{bmatrix} 6 & 0 \\ 0 & 3 \end{bmatrix},$$

then $\mathbf{P}$ would have to be different from above.

In (24), $\mathbf{P}$ is non-singular. Define its inverse as $\mathbf{P}^{-1} = \begin{bmatrix}\mathbf{P}_1^* & \mathbf{P}_2^*\end{bmatrix}$. Then (24) can be rewritten as

$$\begin{aligned}\mathbf{A} &= \mathbf{P}^{-1}\begin{bmatrix}\mathbf{I}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P}^{-1'} = \begin{bmatrix}\mathbf{P}_1^* & \mathbf{P}_2^*\end{bmatrix}\begin{bmatrix}\mathbf{I}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{P}_1^{*'}\\ \mathbf{P}_2^{*'}\end{bmatrix}\\ &= \mathbf{P}_1^*\mathbf{P}_1^{*'}\\ &= \begin{bmatrix}\mathbf{P}_1^* & \mathbf{M}\end{bmatrix}\begin{bmatrix}\mathbf{I}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{P}_1^{*'}\\ \mathbf{N}'\end{bmatrix}\end{aligned}$$

for any $\mathbf{M}$ and $\mathbf{N}$ of order $n \times (n-r)$. Choosing each of $\mathbf{M}$ and $\mathbf{N}$ to have columns that LIN and LIN of those of $\mathbf{P}_1^*$ we can then define

$$\mathbf{P}_0 = \begin{bmatrix}\mathbf{P}_1^* & \mathbf{M}\end{bmatrix}^{-1} \quad \text{and} \quad \mathbf{Q}_0' = \begin{bmatrix}\mathbf{P}_1^* & \mathbf{N}\end{bmatrix}^{-1}$$

and we get

$$\mathbf{P}_0\mathbf{A}\mathbf{Q}_0 = \begin{bmatrix}\mathbf{I}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}.$$

This is an example of an *equivalent canonical form*, a form that can be derived for any matrix, symmetric, non-symmetric or rectangular. Details are shown in Section 7 hereof.

(b) Full rank factorization

Since $\mathbf{U}$ in $\mathbf{A} = \mathbf{UDU}'$ where $\mathbf{U}$ is orthogonal and so non-singular, its columns are linearly independent; and because $\mathbf{A}$ has rank r we will partition $\mathbf{U}$ as $\mathbf{U} = [\mathbf{U}_1 \quad \mathbf{U}_2]$ where $\mathbf{U}_1$ has r columns; and, those columns are linearly independent. Therefore $\mathbf{U}_1$ is $n \times r$ of full column rank. Then from (16)

$$\mathbf{A} = \mathbf{UDU}' = [\mathbf{U}_1 \quad \mathbf{U}_2]\begin{bmatrix}\mathbf{D}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U}_1'\\ \mathbf{U}_2'\end{bmatrix} \tag{25}$$

$$= [\mathbf{U}_1 \quad \mathbf{U}_2]\begin{bmatrix}\sqrt{\mathbf{D}_r} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\sqrt{\mathbf{D}_r} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U}_1'\\ \mathbf{U}_2'\end{bmatrix} \tag{26}$$

and multiplying this out gives

$$\begin{aligned}\mathbf{A}_{n\times n} &= \mathbf{U}_1\sqrt{\mathbf{D}_r}\left(\mathbf{U}_1\sqrt{\mathbf{D}_r}\right)'\\ &= \mathbf{K}_{n\times r}(\mathbf{K}')_{r\times n}\end{aligned} \tag{27}$$

for $\mathbf{K} = \mathbf{U}_1\sqrt{\mathbf{D}_r}$ of order $n \times r$ and full column rank. Result (27) for a symmetric matrix is known as the *full rank factorization*.

Note that in $\mathbf{A} = \mathbf{KK}'$ of (27) the $\mathbf{K}$ does not have to be just $\mathbf{U}_1\sqrt{\mathbf{D}_r}$; it can be post-multiplied by any orthogonal matrix, $\mathbf{Q}$ say, because

$$\mathbf{A} = \mathbf{KK}' = \mathbf{KIK}' = \mathbf{KQQ}'\mathbf{K}' = \mathbf{KQ}(\mathbf{KQ})' .$$

A more general procedure, although not a very useful one, would be to multiply out (26) and get

$$\mathbf{A} = \mathbf{U}_1\,\mathbf{D}_r\,\mathbf{U}_1 = (\mathbf{U}_1\,\mathbf{D}_r)\,\mathbf{U}_1' = \mathbf{V}_1\,\mathbf{U}_1'$$

for $\mathbf{V}_1 = \mathbf{U}_1\,\mathbf{D}_r$, and so $\mathbf{V}_1$ is $\mathrm{n}\times\mathrm{r}$ of full column rank; and $\mathbf{U}_1'$ is $\mathrm{r}\times\mathrm{n}$ of full row rank. Moreover, for any non-singular $\mathbf{S}$ of order r these could be further adapted in the form $\mathbf{A} = (\mathbf{U}_1\mathbf{D}_r\mathbf{S}')(\mathbf{U}_1\mathbf{S}^{-1})'$. Thus there are numerous ways in which $\mathbf{A} = \mathbf{A}'$ can be expressed as a full column rank matrix pre-multiplying a full row rank matrix. Equation (27) is the most useful way.

(c) Diagonal forms

Equation (22) for the particular $\mathbf{W}$ of (21) is

$$\mathbf{A} = \mathbf{UWW}^{-1}\mathbf{DW}'^{-1}\mathbf{W}'\mathbf{U}' .$$

This applies for any non-singular $\mathbf{W}$ of the same order as $\mathbf{A}$. But if that non-singular $\mathbf{W}$ is chosen to be diagonal, then so is $\mathbf{W}^{-1}\mathbf{DW}'^{-1} = \mathbf{D}^*$ say, and so for $\mathbf{V} = \mathbf{UW}$ we have $\mathbf{A} = \mathbf{VD}^*\mathbf{V}'$, where $\mathbf{V}$ is non-singular but not orthogonal (although its columns are). $\mathbf{D}^*$ is diagonal and is often called a *diagonal form* of the matrix $\mathbf{A}$; and for $\mathbf{R} = \mathbf{V}^{-1}$,

$$\mathbf{RAR}' = \mathbf{D}^* \tag{28}$$

for $\mathbf{R}$ non-singular.

(d) Spectral decomposition

For $\mathbf{U}$ partitioned as $\mathbf{U} = [\mathbf{U}_1 \;\; \mathbf{U}_2]$ with $\mathbf{U}_1 = [\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \cdots \;\; \mathbf{u}_r]$,

$$\begin{aligned}
\mathbf{A} = \mathbf{UDU}' &= [\mathbf{U}_1 \;\; \mathbf{U}_2]\begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{U}_1' \\ \mathbf{U}_2' \end{bmatrix} \\
&= \mathbf{U}_1\,\mathbf{D}_r\,\mathbf{U}_1' \\
&= [\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \cdots \;\; \mathbf{u}_r]\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r \end{bmatrix}\begin{bmatrix} \mathbf{u}_1' \\ \mathbf{u}_2' \\ \vdots \\ \mathbf{u}_r' \end{bmatrix} \\
&= \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{u}_i' .
\end{aligned} \tag{29}$$

Result (29) is called the *spectral decomposition* of a symmetric matrix. It is the sum of each non-zero e-root λ_i multiplied by the outer product of the corresponding e-vector, $\mathbf{u}_i$, with itself.

(e) Quadratic forms

From (19) it is clear that

$$\mathbf{x'Ax} = \mathbf{x}' \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{u}_i' \mathbf{x} = \sum_{i=1}^{r} \lambda_i \mathbf{x}' \mathbf{u}_i \mathbf{u}_i' \mathbf{x} = \sum_{i=1}^{r} \lambda_i (\mathbf{u}_i' \mathbf{x})^2 \,. \tag{30}$$

We saw earlier, especially in the example, how $\mathbf{U'AU} = \mathbf{D}$ led to $\mathbf{x'Ax}$ being a sum of squares. And here we see that each of these squares has the form $\left(\sqrt{\lambda_i}\mathbf{u}_i'\mathbf{x}\right)^2$ with a plus or minus sign, depending on whether λ_i is positive or negative. Every symmetric matrix has real numbers as e-roots, but those e-roots can be positive or negative (or zero). Important special cases involving non-negative e-roots are non-negative and positive definite matrices, discussed in the Section (f) that follows:

Example From (18), (19) and (20) we have, for

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \text{ and } \mathbf{U} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & -2 & 0 \\ \sqrt{2} & 1 & -\sqrt{3} \\ \sqrt{2} & 1 & \sqrt{3} \end{bmatrix}, \quad \mathbf{U'AU} = \mathbf{D} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Hence (30) is

$$\Sigma_i \lambda_i (\mathbf{u}_i'\mathbf{x})^2 = 5\left[\sqrt{2}(x_1 + x_2 + x_3)/\sqrt{6}\right]^2 - \left[(-2x_1 + x_2 + x_3)/\sqrt{6}\right]^2 - \left[(-\sqrt{3}x_2 + \sqrt{3}x_3)/\sqrt{6}\right]^2,$$

which is the same as $\mathbf{x'Ax} = \mathbf{y'Dy}$ calculated earlier.

(f) Non-negative definite matrices

-i. Eigen roots are non-negative Suppose $\mathbf{A} = \mathbf{A}'$ is non-negative definite (n.n.d., see text, page 77). Then $\mathbf{x'Ax} \geq 0$ for every vector $\mathbf{x} \neq \mathbf{0}$. Therefore $\mathbf{A} = \mathbf{UDU}'$ gives $\mathbf{x'Ax} = \mathbf{x'UAU'x} = \mathbf{y'Dy}$ for $\mathbf{y} = \mathbf{U'x}$. And, since $\mathbf{U}'$ is orthogonal and so non-singular, any $\mathbf{x} \neq \mathbf{0}$ means that $\mathbf{y} = \mathbf{U'x} \neq \mathbf{0}$. (This is so because $\mathbf{U'x} = \mathbf{0}$ for $\mathbf{x} \neq \mathbf{0}$ would mean that $\mathbf{U}'$ had dependent columns and so $\mathbf{U}'$ would be singular – which it is not.) Therefore $\mathbf{A}$ being n.n.d. means that $\mathbf{y'Dy} \geq 0$ for all $\mathbf{y} = \mathbf{U'x}$. Hence, by letting $\mathbf{y}$ take the value of each column of $\mathbf{I}$ in turn, this means that every diagonal element of $\mathbf{D}$ is non-negative. Thus we have the result that every e-root of n.n.d. matrices is non-negative.

-ii. Positive definite matrices have every e-root positive A special case of this is positive definite (p.d.) matrices. Recall that for $\mathbf{A}$ being p.d., $\mathbf{x'Ax} > 0$ for every $\mathbf{x} \neq \mathbf{0}$. This means the arguments of the preceding paragraph yield the result that for a p.d. matrix every e-root is not just non-negative but is also non-zero; i.e., every e-root is positive. Combining this with the result stated earlier in (ix), that symmetric matrices have rank equal to the number of non-zero e-roots (a result proven in the text, in Section 11.6a), we get the following: for $r(\mathbf{A}) = n$ and

$$\mathbf{A}_{n\times n} = \mathbf{A}'_{n\times n} \text{ being positive definite, every e-root is positive ,}$$

and

$$\mathbf{A}^{-1} \text{ exists .}$$

An important consequence of this for statistics is that for n.n.d. matrices the result in (30) for a quadratic form,

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{r} \lambda_i(\mathbf{u}_i'\mathbf{x})^2 \ ,$$

has only positive values for λ_i. This is important because in statistics every sum of squares can be expressed as a quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$. Then $\mathbf{x}'\mathbf{A}\mathbf{x} = \Sigma_i\lambda_i(\mathbf{u}_i'\mathbf{x})^2$ turns out to be very useful in establishing probability density functions for $\mathbf{x}'\mathbf{A}\mathbf{x}$ when $\mathbf{x}$ represents a vector of normally distributed random variables.

-iii. Congruent canonical form In

$$\mathbf{PAP}' = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

of (24)

$$\mathbf{P} = \mathbf{Q}^{-1} = (\mathbf{UW})^{-1} = \left(\mathbf{U} \begin{bmatrix} \sqrt{\mathbf{D}_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \right)^{-1} .$$

With n.n.d. matrices the diagonal elements of $\mathbf{D}_r$ are positive and so their square roots are real, i.e., are not complex numbers involving $\sqrt{-1}$. Therefore for n.n.d. matrices $\mathbf{P}$ is real, and thus the congruent canonical form of any n.n.d. matrix involves nothing but real matrices. Conversely, for symmetric matrices that are not n.n.d., $\mathbf{P}$ can involve complex numbers.

-iv. Full rank factorization As in (27), a full-rank factorization of a symmetric matrix is

$$\mathbf{A} = \mathbf{KK}' \quad \text{for} \quad \mathbf{K} = \mathbf{U}_1\sqrt{\mathbf{D}_r} \ .$$

For the same reasons that $\mathbf{P}$ of the preceding paragraph is real for n.n.d. matrices, so also is $\mathbf{K}$ in $\mathbf{A} = \mathbf{KK}'$ when $\mathbf{A}$ is n.n.d.

CANONICAL FORMS PART II: ROW OPERATIONS ON ANY MATRIX

4. ROW OPERATORS

(a) Adding a multiple of a row to a row

In dealing with determinants in Chapter 4 we simplified the calculation of a determinant by noting (Section 4.3d, page 95) that the value of a determinant is unaltered by adding to a row a multiple of another row. For example,

$$\begin{vmatrix} 1 & 2 & 4 \\ 3 & 7 & 11 \\ 5 & 8 & 24 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 4 \\ 3+\alpha(1) & 7+\alpha(2) & 11+\alpha(4) \\ 5 & 8 & 24 \end{vmatrix} \tag{31}$$

where to row 2 has been added α times row 1.

This operation, of adding to a row a multiple of another row, can also be applied to a matrix. It is an operation that, perhaps surprisingly, has many uses: and it is not confined to square matrices. Moreover, in contrast to the equality when using the operation on determinants, as illustrated in (31), using the operation on a matrix yields a different matrix. Nevertheless, the two matrices are described as being *equivalent*, denoted by $\cong$. For example,

$$\begin{bmatrix} 1 & 2 & 4 \\ 3 & 7 & 11 \\ 5 & 8 & 24 \\ 1 & -1 & 2 \end{bmatrix} \cong \begin{bmatrix} 1 & 2 & 4 \\ 3+\alpha(1) & 7+\alpha(2) & 11+\alpha(4) \\ 5 & 8 & 24 \\ 1 & -1 & 2 \end{bmatrix} \cong \begin{bmatrix} 1 & 2 & 4 \\ 6 & 13 & 23 \\ 5 & 8 & 24 \\ 1 & -1 & 2 \end{bmatrix},$$

where, for the third matrix, we have taken $\alpha = 3$.

This operation of adding to a row a multiple of another row can be represented as a matrix product by pre-multiplying the initial matrix by a matrix that is an identity matrix amended by having one non-zero value included in its off-diagonal elements. Thus, the middle matrix in the above is

$$\begin{bmatrix} 1 & 2 & 4 \\ 3+\alpha(1) & 7+\alpha(2) & 11+\alpha(4) \\ 5 & 8 & 24 \\ 1 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 7 & 11 \\ 5 & 8 & 24 \\ 1 & -1 & 2 \end{bmatrix}.$$

The matrix with α as its (2,1) element is symbolized as $\mathbf{P}_{21}(\alpha)$ in concert with $\mathbf{P}_{21}(\alpha)\mathbf{A}$ being $\mathbf{A}$ with α

times its row 1 added to its row 2. In general, $\mathbf{P}_{ij}(\alpha)$ is $\mathbf{I}$ amended by having $\alpha \neq 0$ as its (i,j)th element with $i \neq j$. Then $\mathbf{P}_{ij}(\alpha)\mathbf{A}$ adds α times row j of $\mathbf{A}$ to row i of $\mathbf{A}$. Since $\mathbf{P}_{ij}(\alpha)$ is square and essentially

triangular, its determinant is the product of its diagonal elements, i.e., is unity. Thus we find

$$|\mathbf{P}_{ij}(\alpha)| = 1, \qquad [\mathbf{P}_{ij}(\alpha)]^{-1} = \mathbf{P}_{ij}(-\alpha) \qquad \text{and} \qquad [\mathbf{P}_{ij}(\alpha)]' = \mathbf{P}_{ji}(\alpha) .$$

$\mathbf{P}_{ij}(\alpha)$ is one of three matrices that are called *row operator matrices* or just *row operators.* The other two follow.

(b) Interchanging two rows: permutation matrices

A second row operator is the permutation matrix, introduced in Section 6.9. Thus $\mathbf{E}_{24}$ is an *elementary permutation matrix,*

$$\mathbf{E}_{24} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

which is the identity matrix (of order 4 in this case) with its 4th and 2nd rows interchanged. And $\mathbf{E}_{24}\mathbf{A}$ is $\mathbf{A}$ with its 4th and 2nd row interchanged; similarly, $\mathbf{E}_{ij}$ is $\mathbf{I}$ with its i'th and j'th rows interchanged, and $\mathbf{E}_{ij}\mathbf{A}$ is $\mathbf{A}$ with *its* i'th and j'th rows interchanged.

It will also be found that $\mathbf{E}_{ij}$ can be described as $\mathbf{I}$ with its i'th and j'th columns interchanged. This leads to establishing that $\mathbf{E}_{ij}$ is symmetric and orthogonal – and hence non-singular, too. Thus

$$\mathbf{E}_{ij} = \mathbf{E}'_{ij} \qquad \mathbf{E}_{ij}\mathbf{E}'_{ij} = \mathbf{E}^2_{ij} = \mathbf{I} \qquad \text{and} \qquad \mathbf{E}^{-1}_{ij} = \mathbf{E}_{ij} .$$

In contrast to an elementary permutation matrix $\mathbf{E}_{ij}$, is a *permutation matrix,* $\mathbf{P}$ (with no subscripts), which is a product of elementary permutation matrices. Thus $\mathbf{P}$ is $\mathbf{I}$ with its rows resequenced in some manner, as are those of $\mathbf{PA}$. Then, although $\mathbf{P}$ is not symmetric, $\mathbf{P}$ is orthogonal:

$$\mathbf{P}' \neq \mathbf{P} , \qquad \mathbf{PP}' = \mathbf{I} , \qquad \mathbf{P}^{-1} = \mathbf{P}' .$$

And so $\mathbf{P}^{-1}$ is a permutation matrix also.

(c) Multiplying a row by a constant

A third, and final, useful row operation is to multiply a row of a matrix by a constant, θ say. For multiplying the i'th row this is represented by $\mathbf{R}_{ii}(\theta)$, which is $\mathbf{I}$ with its i'th diagonal element being θ rather than unity. Clearly,

$$\mathbf{R}_{ii}(\theta) = [\mathbf{R}_{ii}(\theta)]' \qquad \text{and} \qquad [\mathbf{R}_{ii}(\theta)]^{-1} = \mathbf{R}_{ii}(1/\theta) \ ;$$

and $\mathbf{R}_{ii}(\theta)\mathbf{A}$ is $\mathbf{A}$ with its i'th row multiplied by θ.

5. RANK AND ROW OPERATORS

After multiplying a matrix by any (number) of the three preceding operators the rank of the product matrix is the same as the rank of the initial matrix; i.e., for $\mathbf{Q}$ being any product of any number of $\mathbf{P}_{ij}(\alpha)$, $\mathbf{E}_{ij}$ and $\mathbf{R}_{ij}(\theta)$ matrices, $\mathrm{r}(\mathbf{QA}) = \mathrm{r}(\mathbf{A})$. To illustrate this, suppose in $\mathbf{A}$ that rows $\mathbf{r}_1'$, $\mathbf{r}_2'$ and $\mathbf{r}_3'$ are linearly dependent such that for some scalars a, b and c

$$\mathrm{a}\mathbf{r}_1' + \mathrm{b}\mathbf{r}_2' + \mathrm{c}\mathbf{r}_3' = \mathbf{0} \ . \tag{32}$$

Then in $\mathbf{P}_{12}(\alpha)\mathbf{A}$, for example, the rows are $\mathbf{r}_1' + \alpha\mathbf{r}_2'$, $\mathbf{r}_2'$ and $\mathbf{r}_3'$, therefore, for that matrix,

$$\mathrm{a}(\mathbf{r}_1' + \alpha\mathbf{r}_2') + (\mathrm{b}-\mathrm{a}\alpha)\mathbf{r}_2' + \mathrm{c}\mathbf{r}_3' = \mathrm{a}\mathbf{r}_1' + \mathrm{b}\mathbf{r}_2' + \mathrm{c}\mathbf{r}_3' + \mathrm{a}\alpha\mathbf{r}_2' + \mathrm{a}\alpha\mathbf{r}_2' = 0 \ ;$$

ie., rows of $\mathbf{P}_{12}(\alpha)\mathbf{A}$ are dependent. Hence rank is unaffected. And in $\mathbf{E}_{ij}\mathbf{A}$ the rows are the same row vectors as in $\mathbf{A}$, only in a different sequence. So rank is unaffected there too. And in $\mathbf{R}_{33}(\theta)\mathbf{A}$, for example, the rows are $\mathbf{r}_1'$, $\mathbf{r}_2'$ and $\theta\mathbf{r}_3'$ and so (32) gives, for that product

$$\mathrm{a}\mathbf{r}_1' + \mathrm{b}\mathbf{r}_2' + (\mathrm{c}/\theta)\theta\mathbf{r}_3' = \mathbf{0} \ .$$

So again, rank is unaffected. Thus, in general, row operations on $\mathbf{A}$ yield matrices that have the same rank as $\mathbf{A}$.

Based on this results we can use row operations on any (real) matrix to ascertain its rank. The basic procedure is by means of row operations, to reduce all elements below the (1,1), (2,2), (3,3), $\cdots$ elements to zero. We illustrate with two examples.

(a) Square matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 7 & 11 \\ 5 & 8 & 22 \end{bmatrix}$$

Perform the following row operations:

$$\left.\begin{matrix} \text{to row 2 add } (-3)\text{row 1} \\ \text{to row 3 add } (-5)\text{row 1} \end{matrix}\right\} \quad \text{and get} \quad \mathbf{A} \cong \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & -1 \\ 0 & -2 & 2 \end{bmatrix}$$

and

$$\text{to row 3 add 2(row 2)} \qquad \text{and get} \qquad \mathbf{A} \cong \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}. \tag{33}$$

Since we have carried out only row operations on $\mathbf{A}$, they do not affect rank. Therefore the rank of (33) is the rank of $\mathbf{A}$. Hence we proceed to ascertain the rank of (33); and find that easy to do. Consider its first column: in that column it is impossible for any linear combination of its first two elements (1 and 0) to equal the third element, 0. Therefore no linear combination of rows 1 and 2 of (33) equals row 3. And no multiple of the first row equals the second. Therefore (33) has rank 2, the number of non-null rows: and so $r(\mathbf{A}) = 2$.

This argument extends easily to any matrix, $\mathbf{A}$. Row operations motivated by the values of the elements in column 1 can reduce to zero all elements below the (1, 1) element. Then all those below the (2, 2) element can be similarly changed to zero; and so on until this process can proceed no further. Then, in the resulting matrix, the number of non-null rows is its rank and that is the rank of $\mathbf{A}$.

Sometimes it is necessary to interchange rows in order to have all zeros under each (i, i)th element. For example, consider

$$\mathbf{A}^* = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 2 & 7 \\ 5 & 8 & 22 \end{bmatrix}.$$

Row operations:

$$\left.\begin{array}{l} \text{to row 2 add } (-1)\text{row 1} \\ \text{to row 3 add } (-5)\text{row 2} \end{array}\right\} \qquad \text{gives} \qquad \mathbf{A}^* \cong \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 3 \\ 0 & -2 & 2 \end{bmatrix}$$

and then

$$\text{interchanging row 2 and 3} \qquad \text{gives} \qquad \mathbf{A}^* \cong \begin{bmatrix} 1 & 2 & 4 \\ 0 & -2 & 2 \\ 0 & 0 & 3 \end{bmatrix}. \tag{34}$$

Clearly $r(\mathbf{A}^*) = 3$.

The matrices (33) and (34) are upper triangular matrices (text, p. 14). They are also called *row echelon* forms. In both cases there is a non-decreasing number of zeros in the rows as one goes from the second to third (and for matrices of more than three rows to subsequent) rows. This can also be done for rectangular matrices.

(b) Rectangular matrices

Consider

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 4 & 3 & 5 \\ 3 & 7 & 11 & 8 & 16 \\ 5 & 8 & 22 & 17 & 23 \\ 4 & 9 & 15 & 11 & 28 \end{bmatrix}.$$

We now, for example, represent the row operation "to row 2 add (−3)row 1" more simply as "row 2 − 3(row 1)". Then

$$\left.\begin{matrix} \text{row } 2 - 3(\text{row } 1) \\ \text{row } 3 - 5(\text{row } 1) \\ \text{row } 4 - 4(\text{row } 1) \end{matrix}\right\} \quad \text{gives} \quad \mathbf{B} \cong \begin{bmatrix} 1 & 2 & 4 & 3 & 5 \\ 0 & 1 & -1 & -1 & 1 \\ 0 & -2 & 2 & 2 & -2 \\ 0 & 1 & -1 & -1 & 8 \end{bmatrix};$$

and

$$\left.\begin{matrix} \text{row } 3 + 2(\text{row } 2) \\ \text{row } 4 - \quad \text{row } 2 \end{matrix}\right\} \quad \text{gives} \quad \mathbf{B} \cong \begin{bmatrix} 1 & 2 & 4 & 3 & 5 \\ 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 7 \end{bmatrix}.$$

From this, $r(\mathbf{B}) = 3$. Finally, if non-null rows follow null rows the former can, by interchanging rows, be brought together so that there is a non-decreasing number of zeros in successive rows. For the example, interchanging rows 3 and 4 gives

$$\mathbf{B} \cong \begin{bmatrix} 1 & 2 & 4 & 3 & 5 \\ 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{35}$$

These examples demonstrate that using a series of row operations on a matrix can lead to a matrix whose rank is self-evident and that is the rank of the original matrix. For matrices of any but small dimension the arithmetic can be very tedious. Nevertheless seeing the details for small classroom matrices does illustrate and emphasize that there is a way in which we can find the rank of any numerical matrix. But in practice we would actually do it by using computer software designed for the purpose. So we need go no further with this arithmetic.

Moreover, in many situations we deal with matrices that can be expressed in an algebraic form. And through understanding the concept of linear independence and knowing that rank is the number of LIN rows (or columns), algebraic forms often reveal the rank of a matrix without any arithmetic at all.

The following examples illustrate this

$$\begin{bmatrix} a & b & c & d \\ x & y & z & w \\ a+x & b+y & c+z & d+w \end{bmatrix} \text{has rank 2.} \qquad \begin{bmatrix} a & b & c & d \\ a+1 & b+1 & c+1 & d+1 \end{bmatrix} \text{has rank 2.}$$

$$\begin{bmatrix} a & b & c & d \\ a^2 & b^2 & c^2 & d^2 \end{bmatrix} \text{has rank 2.} \qquad \begin{bmatrix} a & b & c & d \\ a^2 & b^2 & c^2 & d^2 \\ a(a+1) & b(b+1) & c(c+1) & d(d+1) \end{bmatrix} \text{has rank 2.}$$

$$\begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X} \end{bmatrix} \text{has rank twice the rank of } \mathbf{X}. \qquad \begin{bmatrix} \mathbf{I}_n & \mathbf{K} \end{bmatrix} \text{has rank n.}$$

Hence, in practice, the rank of a matrix can be obtained either from computer software, or from knowing the algebraic form of the matrix, or from both.

6. DIAGONAL FORMS

(a) An example

For the preceding example **B**, of order 4×5 (more columns than rows), we arrived at **B** equivalent to the matrix in (35). In that matrix consider the square matrix of the first 4 columns (because **B** has 4 rows) and note that it is an upper triangular matrix,

$$\mathbf{T} = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{36}$$

albeit with two of its diagonal elements being zero. Thus we can write (35) as

$$\mathbf{B} \cong [\mathbf{T} \quad \mathbf{S}]$$

for **T** square and upper triangular. This has been achieved with row operations. Since each of those operations can be represented either by a matrix $\mathbf{P}_{ij}(\alpha)$, or by a permutation matrix $\mathbf{E}_{ij}$, the totality of those operations can be represented by the product of the individual $\mathbf{P}_{ij}(\alpha)$ and $\mathbf{E}_{ij}$ matrices. Call that product **P**, with no subscript. Then

$$\mathbf{PB} = [\mathbf{T} \quad \mathbf{S}] \,. \tag{37}$$

We now show how a further set of row operations can yield a matrix (equivalent to **B**) having its only non-zero elements as elements (1, 1), (2, 2), (3, 3), $\cdots$. We refer to these as "diagonal" elements even though **B** is not square. The procedure is as follows.

In keeping with (37) denote the matrix in (35) as **PB** and transpose it:

$$(\mathbf{PB})' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 4 & -1 & 0 & 0 \\ 3 & -1 & 0 & 0 \\ 5 & 1 & 7 & 0 \end{bmatrix}. \tag{38}$$

If, as here, interchanging a pair of rows can change a zero "diagonal" element to be non-zero, make that interchange; or several such interchanges as are necessary to change more than one zero "diagonal" element to be non-zero. Call the product of corresponding row operations $\mathbf{Q}_0'$. Then for the example, on (38)

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 4 & -1 & 0 & 0 \\ 3 & -1 & 0 & 0 \\ 5 & 1 & 7 & 0 \end{bmatrix} = \mathbf{Q}_0(\mathbf{PB})' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 5 & 1 & 7 & 0 \\ 3 & -1 & 0 & 0 \\ 4 & -1 & 0 & 0 \end{bmatrix}. \tag{39}$$

Then on $(\mathbf{Q}_0\mathbf{PB})'$ the row operations

$$\left.\begin{array}{l} \text{row } 2 - 2(\text{row } 1) \\ \text{row } 3 - 5(\text{row } 1) \\ \text{row } 4 - 3(\text{row } 1) \\ \text{row } 5 - 4(\text{row } 1) \end{array}\right\} \quad \text{gives} \quad \mathbf{Q}_0(\mathbf{PB})' \cong \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 7 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}.$$

Next

$$\left.\begin{array}{l} \text{row } 3 - \text{row } 2 \\ \text{row } 4 + \text{row } 2 \\ \text{row } 5 + \text{row } 2 \end{array}\right\} \quad \text{gives} \quad \mathbf{Q}_0(\mathbf{PB})' \cong \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{40}$$

As with deriving **PB** itself, each of the preceding row operations can be represented by a row operator matrix, and the product of these matrices and $\mathbf{Q}_0$ will represent the totality of row operations on $(\mathbf{PB})'$. Call that product $\mathbf{Q}'$. Then (40) is

$$\mathbf{Q}'(\mathbf{PB})' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{D}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{5 \times 4}$$

or, on transposing

$$\mathbf{PBQ} = \begin{bmatrix} \mathbf{D}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{4 \times 5} \qquad \text{where} \qquad \mathbf{D}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix} \tag{41}$$

and the three null matrices in (41) are of orders sufficient to make the **PBQ** matrix 4×5.

Because $\mathbf{D}_3$ is diagonal, (41) is called a *diagonal form* of **B**. It is "a" diagonal form, not "the" diagonal form, because it is not unique. For example, additional permutation matrices included in **P** and/or $\mathbf{Q}'$ could lead to a $\mathbf{D}_3$ different from (41). But for this **B** there will always be a $\mathbf{D}_3$, a 3×3 diagonal matrix with non-zero diagonal elements. This is so because **B** has rank 3.

The important feature of **P** and **Q** in (41) is not their numerical value but the fact that they always exist. Although their numeric values are seldom of use or interest, those values can, of course, be obtained. This is done by performing row operations on **I**. For **PB** being the matrix in (35), **P** is obtained by carrying out on $\mathbf{I}_4$ (because **B** has 4 rows) the five row operations on **B** shown leading up to (35). Likewise $\mathbf{Q}'$ is obtained by carrying out on $\mathbf{I}_5$ [because $(\mathbf{PB})'$ has 5 columns] the eight row operations on $(\mathbf{PB})'$ shown between (35) and (40). The reader might like to do these calculations and check that the resulting **P** and **Q** give $(\mathbf{PB})'$ of (38) and **PBQ** of (41).

(b) The general case

We consider **A** of order $p \times q$ and rank r. Then, if $p \leq q$ as in the example **B**, we can have, akin to (37),

$$\mathbf{PA} = [\mathbf{T} \quad \mathbf{S}]_{p \times q} \tag{42}$$

where **T** is upper triangular. Transposing this gives

$$(\mathbf{PA})' = \begin{bmatrix} \mathbf{T}' \\ \mathbf{S}' \end{bmatrix}. \tag{43}$$

This is a matrix of more rows that columns, just like $(\mathbf{PB})'$ of (38). Therefore row operations on $(\mathbf{PA})'$ can reduce to zero all elements below its "diagonal", just as illustrated leading up to (40). Thus on denoting these row operations by $\mathbf{Q}'$ we reduce $(\mathbf{PA})'$ to have the same form as (40), initially denoting it as

$$\mathbf{Q}'(\mathbf{PA})' = \begin{bmatrix} \mathbf{T}^* \\ \mathbf{0} \end{bmatrix}, \tag{44}$$

for $\mathbf{T}^*$ being square. In getting to here, just as illustrated in going from (38) to (39) to (40), $\mathbf{T}^*$ has non-zero elements only in its diagonal. There will be $r = r(\mathbf{A})$ such elements, because only row operations have been used to get $\mathbf{Q}'(\mathbf{PA})'$ and so

$$r(\mathbf{A}) = r[\mathbf{Q}'(\mathbf{PA})'] = r(\mathbf{T}^*) \ , \tag{45}$$

this latter equality coming directly from the occurrence of the null matrix in (44). Finally, note that we purposefully use the row operations to bring all the non-zero diagonal elements of $\mathbf{T}^*$ to be the first r diagonal elements. Therefore, $\mathbf{Q}'(\mathbf{PA})'$ can be written in partitioned form as

$$\mathbf{Q}'(\mathbf{PA})' = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{q \times p} \tag{46}$$

and transposing gives

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{p \times q} . \tag{47}$$

This is a diagonal form of $\mathbf{A}$.

When $p \geq q$ we deal with $\mathbf{A}'$ in similar manner, only using the symbols $\mathbf{Q}'$ and $\mathbf{P}$ in slightly different positions, but still representing products of row operators. Thus we start with

$$\mathbf{Q}'\mathbf{A}' = [\mathbf{T} \quad \mathbf{S}]_{q \times p} \tag{48}$$

and then use a $\mathbf{P}$ such that

$$\mathbf{P}(\mathbf{Q}'\mathbf{A}')' = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{p \times q} . \tag{49}$$

These two procedures are operating on $(\mathbf{A}')_{q \times p}$ with $q \leq p$ just like (42) and (46) are operating on $\mathbf{A}_{p \times q}$ with $p \leq q$; and in (48) and (49) we use symbols $\mathbf{Q}'$ and $\mathbf{P}$ where in (42) and (46) we used $\mathbf{P}$ and $\mathbf{Q}'$,

respectively. Then, simplifying (49) gives

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{p \times q} , \tag{50}$$

the same form as (47). Thus, no matter what p and q are, we have (50) as a diagonal form of $\mathbf{A}_{p \times q}$.

(c) Conclusion

The prime importance of (50) is that it applies for *any* matrix (of real numbers) $\mathbf{A}$. Thus for any

$$\mathbf{A}_{p \times q} \quad \text{of} \quad \text{rank r} ,$$

there is *always* a diagonal form

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{p \times q} \tag{51}$$

where

> $\mathbf{P}$ and $\mathbf{Q}$ are not unique for given $\mathbf{A}$,
>
> $\mathbf{P}$ and $\mathbf{Q}'$ are products of row operator matrices ,
>
> $\mathbf{P}^{-1}$ and $\mathbf{Q}^{-1}$ exist

and

> $\mathbf{D}_r$ is diagonal, with r non-zero diagonal elements .

7. EQUIVALENT CANONICAL FORMS

Suppose (51) is pre-multiplied by the non-singular matrix

$$\mathbf{D}^* = \begin{bmatrix} \mathbf{D}_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-r} \end{bmatrix}, \quad \text{giving} \quad \mathbf{D}^*\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{p \times q} . \tag{52}$$

$\mathbf{D}_r$ is simply a product of row operators of the form $\mathbf{R}_{ii}(\lambda)$. Therefore $\mathbf{D}^*\mathbf{P}$ is a product of row operators, and so on re-defining the symbol $\mathbf{P}$ to be the $\mathbf{D}^*\mathbf{P}$ in (52) we get

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{p \times q} , \tag{53}$$

with $\mathbf{P}$ and $\mathbf{Q}$ having the same properties as in (51). Result (53) is known as an *equivalent canonical form* of $\mathbf{A}$. It is very useful. For non-symmetric (including rectangular) matrices it is the counterpart of the congruent canonical form (24) for symmetric matrices.

As stated for $\mathbf{P}$ and $\mathbf{Q}$ in (51), so also in (53): the important feature of $\mathbf{P}$ and $\mathbf{Q}$ is that they always exist; and they have inverses. It is their existence that is so useful and important. Their numerical values are seldom needed. The usefulness of their existence, without having to know their explicit values, is demonstrated in the three important results that follow.

8. USING THE EQUIVALENT CANONICAL FORM

(a) Full rank factorization

Because $\mathbf{P}$ and $\mathbf{Q}$ in the equivalent canonical form

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

are non-singular,

$$\mathbf{A} = \mathbf{P}^{-1}\begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{Q}^{-1} .$$

Define

$$\mathbf{P}^{-1} = [\mathbf{K} \quad \mathbf{K}_1] \qquad \text{and} \qquad \mathbf{Q}^{-1} = \begin{bmatrix} \mathbf{L} \\ \mathbf{L}_1 \end{bmatrix},$$

where $\mathbf{K}$ has r columns and $\mathbf{L}$ has r rows. Then

$$\mathbf{A}_{p \times q} = [\mathbf{K} \quad \mathbf{K}_1]\begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{L} \\ \mathbf{L}_1 \end{bmatrix} = \mathbf{KL} .$$

Because $\mathbf{K}$ consists of r columns of $\mathbf{P}^{-1}$, which is a non-singular matrix, the columns of $\mathbf{K}$ are LIN; i.e., $\mathbf{K}_{p \times r}$ has full column rank. Similarly $\mathbf{L}$ has r rows of $\mathbf{Q}^{-1}$ that are therefore LIN, and so $\mathbf{L}_{r \times q}$ has full row rank. Thus for

$$\mathbf{A}_{p \times q} \text{ of rank r}, \quad \text{we have } \mathbf{A}_{p \times q} = \mathbf{K}_{p \times r}\mathbf{L}_{r \times q} \tag{54}$$

where $\mathbf{K}$ has full column rank and $\mathbf{L}$ has full row rank. This is known as the *full rank factorization.* Recall from (27) that when $\mathbf{A}$ is a symmetric matrix its full rank factorization is

$$\mathbf{A} = \mathbf{K}_{p \times r}(\mathbf{K}')_{r \times q} = \mathbf{A}' .$$

The existence of this form was established in (27) based on the canonical form under orthogonal similarity of $\mathbf{A} = \mathbf{A}'$. But, as shown there, the result is more general than being specifically tied to the canonical form under orthogonal similarity.

One reason for full rank factorization being useful is that through involving full row rank and full column rank matrices, use can be made of the result that for $\mathbf{K}$ of full column rank, $(\mathbf{K}'\mathbf{K})^{-1}$ exists. This is so because we can write $\mathbf{K}$ (or perhaps $\mathbf{PK}$ where $\mathbf{P}$ is a permutation matrix) as

$$\mathbf{K}_{p \times r} = \begin{bmatrix} \mathbf{M} \\ \mathbf{SM} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{S} \end{bmatrix}\mathbf{M} \tag{55}$$

for $\mathbf{M}$ non-singular of rank r and for some $\mathbf{S}$. Then

$$\mathbf{K'K} = \mathbf{M'(I+S'S)\,M} \;.$$

Now $\mathbf{M}^{-1}$ exists; and so does $(\mathbf{I+S'S})^{-1}$, because $\mathbf{x'(I+S'S)\,x} = \mathbf{x'x} + (\mathbf{Sx})'(\mathbf{Sx})$ can never be zero or negative for $\mathbf{x} \neq \mathbf{0}$. Hence $\mathbf{I+S'S}$ is positive definite, and so has an inverse. Therefore

$$(\mathbf{K'K})^{-1} = \left[\mathbf{M'(I+S'S)\,M}\right]^{-1} = \mathbf{M}^{-1}(\mathbf{I+S'S})^{-1}\,\mathbf{M'}^{-1}$$

exists. When $\mathbf{PK}$ is used in place of $\mathbf{K}$ in (55) we get the inverse

$$(\mathbf{K'P'PK})^{-1} = \left[\mathbf{M'(I+S'S)\,M}\right]^{-1} = (\mathbf{K'K})^{-1}$$

because $\mathbf{P}$ is orthogonal.

(b) Non-singular matrices

In the equivalent canonical form

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{p \times q}$$

the dimensions of the null matrices depend upon r, p and q. When $\mathbf{A}$ is non-singular, r = p = q and the preceding equation is

$$\mathbf{PAQ} = \mathbf{I} \,,$$

i.e.,

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{Q}^{-1} \;. \tag{56}$$

Now $\mathbf{P}$ is a product of a series of row operators, each of which is one of either $\mathbf{P}_{ij}$, $\mathbf{E}_{ij}$ or $\mathbf{R}_{ii}(\lambda)$. Therefore $\mathbf{P}^{-1}$ is a product of the corresponding inverses

$$\left[\mathbf{P}_{ij}(\alpha)\right]^{-1} = \mathbf{P}_{ij}(-\alpha), \qquad \mathbf{E}_{ij}^{-1} = \mathbf{E}_{ij} \qquad \text{and} \qquad \left[\mathbf{R}_{ii}(\lambda)\right]^{-1} = \mathbf{R}_{ii}(1/\lambda) \;.$$

But these inverses are themselves seen to be row operators and so, therefore, is $\mathbf{P}^{-1}$. The same is true of $\mathbf{Q}^{-1}$. Hence in (56), $\mathbf{A}$ is a product of one row operator matrix and the transpose of another.

(c) Rank of a product matrix

One of the most useful results concerning rank involves inequality statements about the rank of a product of two matrices. It is simply stated, as follows.

Theorem. $r_{AB} \leq$ the lesser of r_A and r_B.

Proof. For $r_A = r$ we know that non-singular $\mathbf{P}$ and $\mathbf{Q}$ exist such that

$$\mathbf{PAQ} = \mathbf{C} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and equivalently} \quad \mathbf{PA} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}^{-1} .$$

Post-multiply both sides of the second equation by $\mathbf{B}$, and for $\mathbf{B}$ having n columns partition $\mathbf{Q}^{-1}\mathbf{B}$ as $\mathbf{Q}^{-1}\mathbf{B} = \begin{bmatrix} \mathbf{T}_{r \times n} \\ \mathbf{S} \end{bmatrix}$ so that

$$\mathbf{PAB} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}^{-1}\mathbf{B} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{T}_{r \times n} \\ \mathbf{S} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{r \times n} \\ \mathbf{0} \end{bmatrix} .$$

Then, because $\mathbf{P}$ is a product of row operators,

$$r_{AB} = r_{PAB} = r\begin{bmatrix} \mathbf{T}_{r \times n} \\ \mathbf{0} \end{bmatrix} \leq r \text{ because } \mathbf{T} \text{ has r rows ;}$$

i.e., $r_{AB} \leq r_A$, since $r = r_A$. Similar reasoning, based on the canonical form of $\mathbf{B}$, gives $r_{AB} \leq r_B$. QED

Even though this theorem does not involve an equality statement, it can often be applied twice to yield inequalities of the form rank $\leq$ m and m $\leq$ rank and hence rank = m. The following lemmas illustrate this use. The lemmas themselves find widespread use through matrix algebra, and the proofs of them illustrate how useful the theorem is.

Lemma 1. Multiplication by a non-singular matrix does not alter rank.

Proof. Let $\mathbf{Q} = \mathbf{HA}$ for $\mathbf{H}$ non-singular. Then, by the theorem $r_Q \leq r_A$. But $\mathbf{A} = \mathbf{H}^{-1}\mathbf{Q}$ so that the theorem also gives $r_A \leq r_Q$. Hence $r_Q = r_A$. QED

This lemma is used repeatedly where problems of rank are concerned.

Lemma 2. For arbitrary $\mathbf{z}$ the maximum number of LIN vectors $\mathbf{Bz}$ is r_B.

Proof. Let $\mathbf{Z}$ be a non-singular matrix of the same order as $\mathbf{z}$. Then, by Lemma 1, $r_{BZ} = r_B$ and so the maximum number of LIN vectors $\mathbf{Bz}$, i.e., the maximum number of LIN columns in $\mathbf{BZ}$, is r_B. QED

This lemma plays an important role in solving linear equations (see Section 10.4).

Lemma 3. If $\mathbf{AGA} = \mathbf{A}$, then $r_{GA} = r_A$.

Proof. Applying the theorem directly to $\mathbf{GA}$ gives $r_{GA} \leq r_A$, and applied to $\mathbf{A} = \mathbf{AGA}$ it gives $r_A \leq r_{GA}$. Therefore $r_{GA} = r_A$. QED

Matrices $\mathbf{G}$ of this form are called generalized inverses and are the subject of Chapter 8. They are used repeatedly in solving equations, as discussed in Chapter 9.

9. THE Q – R FACTORIZATION

A useful factorization in statistics that is known as the Q – R factorization is $\mathbf{A} = \mathbf{QR}$ where for square $\mathbf{A}$ the $\mathbf{Q}$ is orthogonal and for $\mathbf{A}$ of more rows than columns $\mathbf{Q}$ has orthonormal columns; and in both cases $\mathbf{R}$ is square and upper triangular.

One way of deriving this factorization is based on Householder matrices, $\mathbf{H} = \mathbf{I} - 2\mathbf{hh}'$ where $\mathbf{h}'\mathbf{h} = 1$, as described in Section 3.4b(iii) of the text. $\mathbf{H}$ of this form is symmetric and orthogonal:

$$\mathbf{H} = \mathbf{H}' = \mathbf{H}^{-1} \quad \text{and} \quad \mathbf{HH}' = \mathbf{H}'\mathbf{H} = \mathbf{H}^2 = \mathbf{I}\,.$$

A particular feature of $\mathbf{H}$ is that for a vector $\mathbf{x}$ of order n it is always possible to find an $\mathbf{H}$ of order n such that

$$\mathbf{Hx} = \begin{bmatrix} \lambda \\ \mathbf{0} \end{bmatrix}. \tag{57}$$

This is spelled out on page 73 of the text. When $\mathbf{H}$ and $\mathbf{x}$ are of the form (57) we speak of "$\mathbf{H}$ for $\mathbf{x}$". It is the basis of developing the QR transformation by deriving first, an $\mathbf{H}_1$ for $\mathbf{x}_1$ as the first column of $\mathbf{A}$. Then a second $\mathbf{H}$ is developed, $\mathbf{H}_2$, for $\mathbf{x}_2$ being the second column of $\mathbf{H}_1\mathbf{A}$ excluding its first element. Thirdly, an $\mathbf{H}_3$ is calculated for $\mathbf{x}_3$ being the third column (excluding the first two elements) of

$$\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix} \mathbf{H}_1\mathbf{A}\,. \tag{58}$$

And fourthly is $\mathbf{H}_4$ for $\mathbf{x}_4$ being the fourth column (excluding the first three elements) of

$$\begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_3 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix} \mathbf{H}_1\mathbf{A} \tag{59}$$

and so on. At each stage, $\mathbf{H}_1$, and the matrix products pre-multiplying $\mathbf{A}$ in (58), (59) and so on are orthogonal. For example, for $\mathbf{H}^*\mathbf{A}$ being (58)

$$\mathbf{H}^*\mathbf{H}^{*\prime} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix} \mathbf{H}_1 \left\{ \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix} \mathbf{H}_1 \right\}' = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix} \mathbf{H}_1\mathbf{H}_1' \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2' \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix} \mathbf{I} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2' \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2\mathbf{H}_2' \end{bmatrix} = \mathbf{I}\,.$$

We demonstrate, for square matrices, by using the example of Exercise 22 on pages 82-3 of the text.

(a) Example: a square matrix

Take
$$\mathbf{A} = \begin{bmatrix} 1 & -3 & 88 \\ 2 & 4 & 1 \\ 2 & 5 & 0 \end{bmatrix}.$$

$\mathbf{H}$ for the first column of $\mathbf{A}$,

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \quad \text{is} \quad \mathbf{H}_1 = \tfrac{1}{3}\begin{bmatrix} -1 & -2 & -2 \\ -2 & 2 & -1 \\ -2 & -1 & 2 \end{bmatrix} \quad \text{with} \quad \mathbf{H}_1\mathbf{x}_1 = \begin{bmatrix} -3 \\ 0 \\ 0 \end{bmatrix};$$

and
$$\mathbf{H}_1\mathbf{A} = \begin{bmatrix} -9 & -5 & -30 \\ 0 & 3 & -58 \\ 0 & 4 & -59 \end{bmatrix}.$$

We now find $\mathbf{H}_2$ for

$$\mathbf{x}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}. \text{ It is } \mathbf{H}_2 = \tfrac{1}{5}\begin{bmatrix} -3 & -4 \\ -4 & 3 \end{bmatrix} \quad \text{with} \quad \mathbf{H}_2\mathbf{x}_2 = \begin{bmatrix} -5 \\ \mathbf{0} \end{bmatrix}.$$

And then

$$\mathbf{H}^* = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix}\mathbf{H}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{-3}{5} & \frac{-4}{5} \\ 0 & \frac{-4}{5} & \frac{3}{5} \end{bmatrix}\mathbf{H}_1 = \tfrac{1}{15}\begin{bmatrix} -5 & -10 & -10 \\ 14 & -2 & -5 \\ 2 & -11 & 10 \end{bmatrix},$$

Therefore

$$\mathbf{A} = \mathbf{Q}_1\mathbf{R} \quad \text{with} \quad \mathbf{R} = \begin{bmatrix} -3 & -5 & -30 \\ 0 & -5 & 82 \\ 0 & 0 & 11 \end{bmatrix}$$

and

$$\mathbf{Q}_1 = \left(\mathbf{H}^*\right)^{-1} = \mathbf{H}_1^{-1}\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2^{-1} \end{bmatrix} = \mathbf{H}_1\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix} = \mathbf{H}^{*\prime} = \tfrac{1}{15}\begin{bmatrix} -5 & 14 & 2 \\ -10 & -2 & -11 \\ -10 & -5 & 10 \end{bmatrix}.$$

For square $\mathbf{A}$, non-singular, as here, $\mathbf{R}$ is square and upper triangular because of the nature of the successive $\mathbf{H}$-matrices, each of which is of the form that satisfies (57). If $\mathbf{A}$ is singular, $\mathbf{R}$ will be upper triangular but singular. And so we will have one or more null rows. And for $\mathbf{A}$ of order n there will be n − 1 $\mathbf{H}$-matrices.

If one wants the negative diagonal elements of $\mathbf{R}$ to be positive it is easy to introduce a diagonal matrix $\boldsymbol{\Delta}$, say, that has terms +1 and −1 corresponding, respectively, to positive and negative diagonal elements in $\mathbf{R}$. Then $\boldsymbol{\Delta}\mathbf{H}^*\mathbf{A}$ is $\mathbf{R}$ with rows corresponding to negative diagonal elements of $\mathbf{R}$

multiplied by (–1). And $\mathbf{\Delta H^*}$ is orthogonal and so is $\mathbf{Q} = (\mathbf{\Delta H^*})^{-1} = \mathbf{H^{*\prime}\Delta}$ and $\mathbf{Q_1 R} = \mathbf{H^{*\prime}(\Delta R)}$.

(b) Rectangular matrices: more rows than columns

Suppose $\mathbf{A}$ is p + q with p > q. All of the preceding development still holds except that we will have an $\mathbf{H^*}$ such that

$$\mathbf{H^* A} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}$$

where $\mathbf{R}$ is upper triangular.

(c) Inverting A = QR

When $\mathbf{A}$ is non-singular the $\mathbf{A} = \mathbf{QR}$ factorization is very useful computationally for inverting $\mathbf{A}$ because then

$$\mathbf{A}^{-1} = \mathbf{R}^{-1}\mathbf{Q}^{-1} = \mathbf{R}^{-1}\mathbf{H}^* ,$$

and with $\mathbf{R}$ being upper triangular its inverse is easy to calculate – with simple algorithms (having good numerical accuracy) for doing so.